Database Integration for Data and Metadata Management
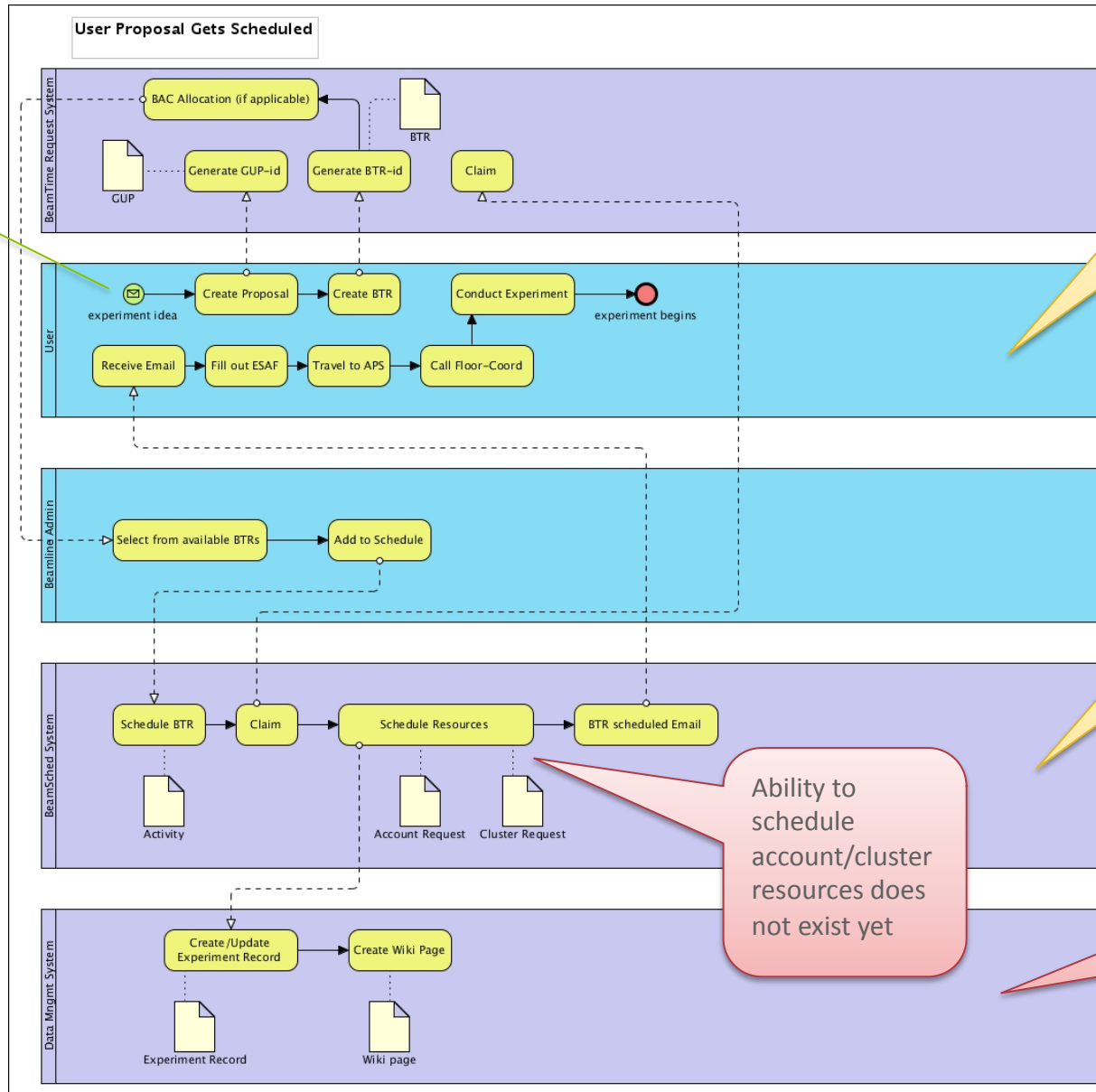
Claude Saunders

Software Services Group

for Tomography Upgrade Project

Tuesday, May 10, 2011

U.S. DEPARTMENT OF **ENERGY**

Process Diagrams (1)

- 4 Process Diagrams outline how database systems will interact with typical tomography upgrade workflow.

   1. User Proposal Gets Scheduled
   2. Account Request Handled
   3. User Conducts Experiment
   4. User Retrieves Data Files

Beamline Scheduling System

- Exists already!
- Contains all fields for Tomo Script User and Facility tab
  - Plus more – pretty much all scheduling/GUP/ User data

- Need to enhance "Activity" dialog to allow for making account/cluster requests
  - Already have resource request tables built into database schema
- Need to extend email queuing mechanism to queue account/cluster requests
  - account provisioning can feed off this queue

Edit Activity: ✕

Category: [Beamtime Request ▼]

Find: [_____]   [Prev] [Next]

Prop	PI	req/grt	sch	bl
22295	Faber	6/3 shifts	6/6 shifts	1st
22329	Kotha	9/9 shifts	12/12 shift:	1st
11370	Stock	9/9 shifts	6/6 shifts	1st
22782	La Riviere	18/9 shifts	9/9 shifts	1st
22737	Lee	9/9 shifts	0/0 shifts	3rd
24201	Wen	4/0 shifts	0/0 shifts	1st

Proposal: General User   BTR: 54119

Status: Claimed by 2 - BM

Title: Microstructural Characterization of Biomorphic Carbon/Copper Composites

PI: Katherine Faber

Inst: Northwestern University

ESAF ID: [76125]   Access: local

Comment: [_____]

Start Time: [10/13/2010] 🗓 08 : 00 ▲▼   Color:

End Time: [10/15/2010] 🗓 08 : 00 ▲▼   ■

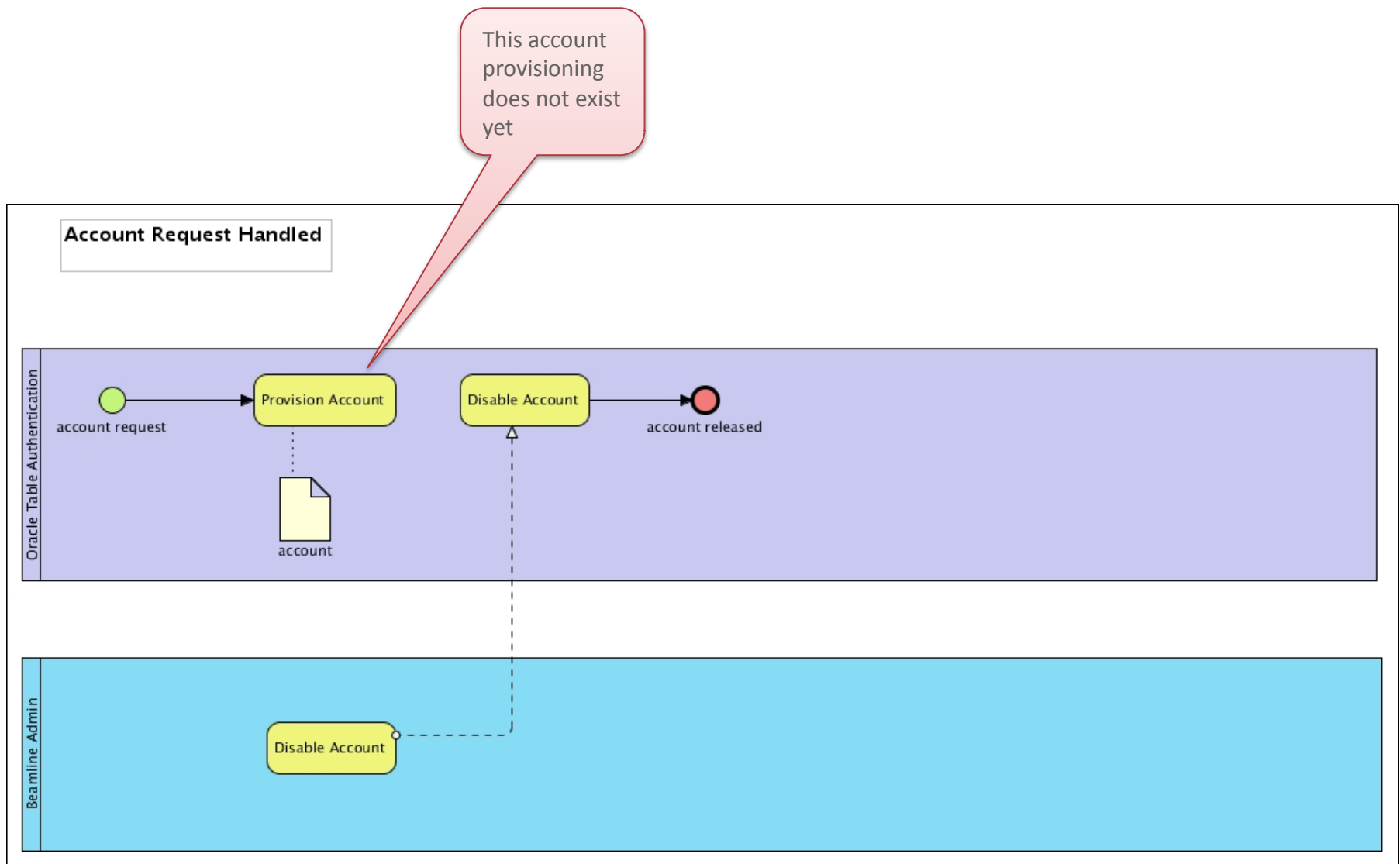Time Delta: 0 hours   ☐ unused
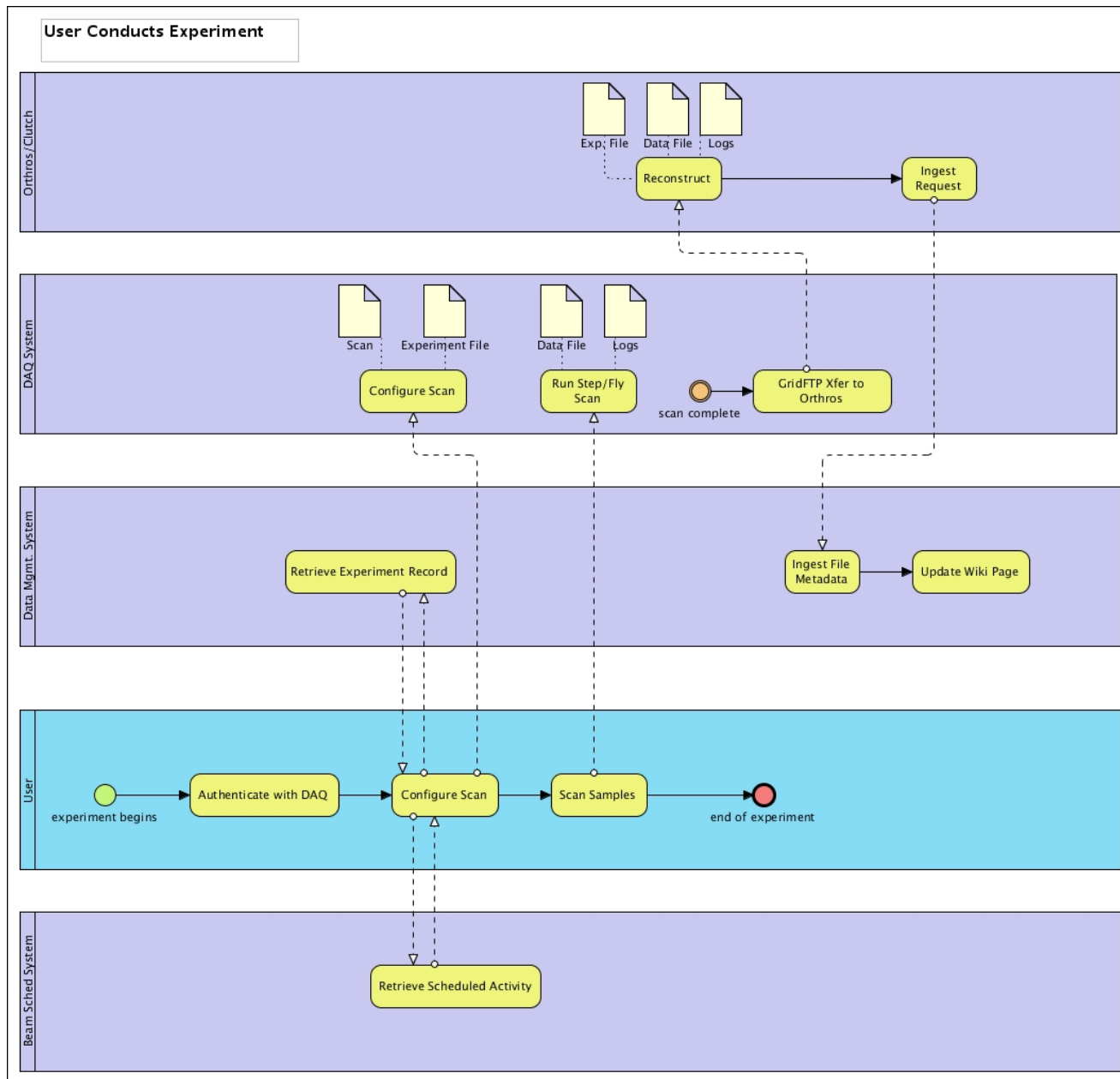
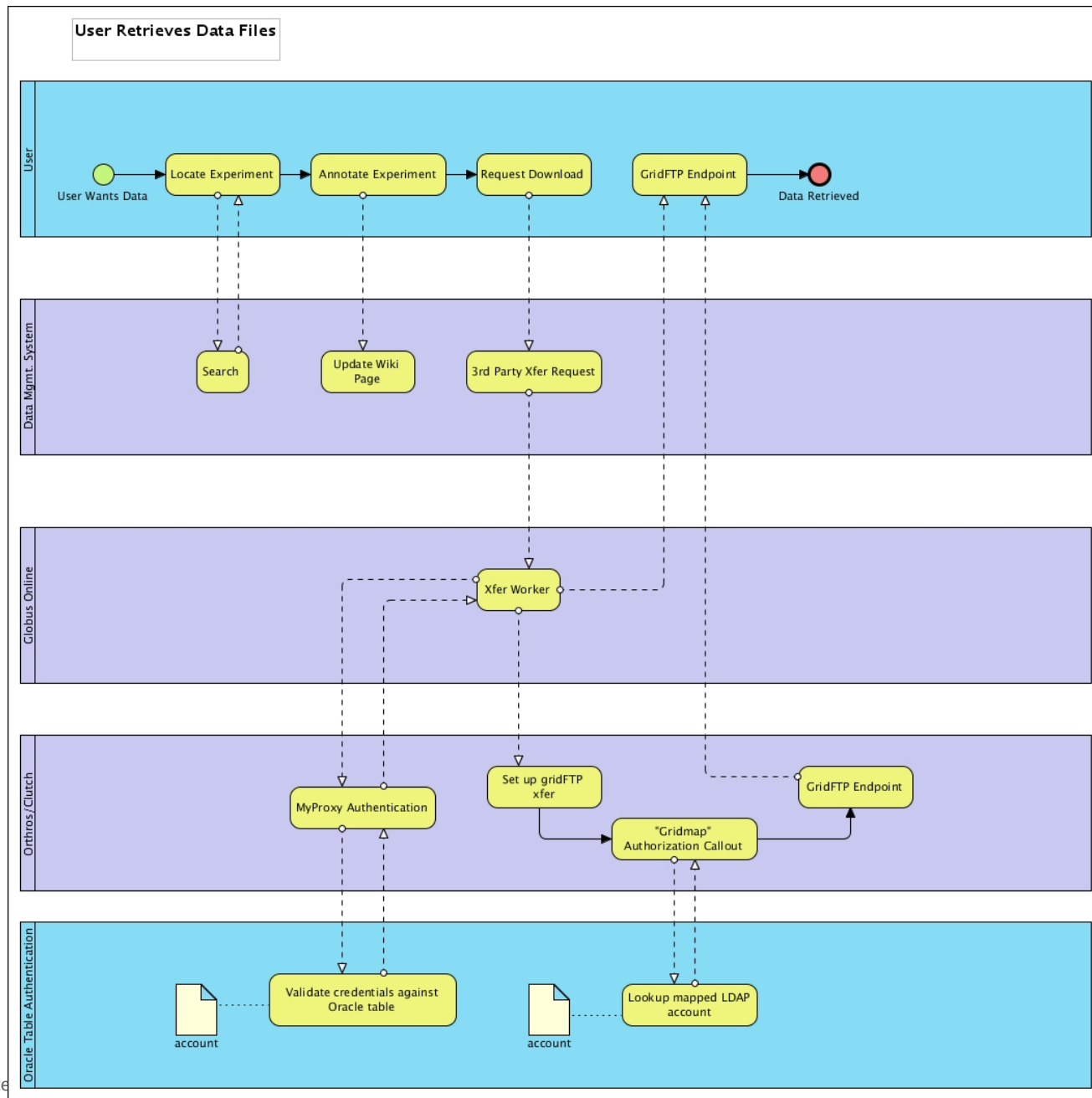[Activity] [Activity Email]

[Apply] [Delete]

Data Management System

- Does not exist

- Includes (potentially):
  - Experiment/investigation/sample database
  - Browser interfaces for human interaction (find files and request download)
  - Services for programmatic interaction
  - File ingestion for scanning data/experiment/logs to index them

- Ideally, would be coupled with a Beamline Notebook (in scope?...)

- Beamline notebook (wiki) tightly coupled to Data Management System
  - Presents wiki-style page summarizing experiment
    - With links to access data files, provenance, logs, etc
  - Automatically created based on records kept by Data Management System
  - Scientist can annotate manually

User Retrieves Data Files

User
- User Wants Data
- Locate Experiment
- Annotate Experiment
- Request Download
- GridFTP Endpoint
- Data Retrieved

Data Mgmt. System
- Search
- Update Wiki Page
- 3rd Party Xfer Request

Globus Online
- Xfer Worker

Orthros / Clutch
- MyProxy Authentication
- Set up gridFTP xfer
- "Gridmap" Authorization Callout
- GridFTP Endpoint

Oracle Table Authentication
- Validate credentials against Oracle table
- account
- Lookup mapped LDAP account
- account

Database Inte

Candidate Data Management Systems

- ICAT from ISIS/ILL/Diamond (UK)
  - http://code.google.com/p/icatproject
  - Got a local instance of ICAT running on my desktop system last year
  - Heavyweight architecture (uses just about every J2EE technology in Glassfish)
  - Found web services interface difficult to understand and work with

- myTARDIS (multi-institution collaboration, used at Australian Synchrotron)
  - http://code.google.com/p/mytardis
  - Python/Django based
  - Debby Quock is getting local instance up to report on capabilities

- We can always roll our own, or other options may come to our attention (MCS?)

- Wiki
  - MediaWiki has REST-based API for programmatic interaction
  - PHP-based